TITLE OF THE INVENTION

A PROCESS TO EXTRACT REGIONS OF HOMOGENEOUS COLOR IN A DIGITAL PICTURE

CROSS REFERENCE TO RELATED APPLICATIONS

This is a continuation of provisional U.S. Patent Application Serial No. 60/118,192 filed February 1, 1999, now abandoned.

BACKGROUND OF THE INVENTION

The present invention relates to video data processing, and more particularly to a process for extracting regions of homogeneous color in a digital picture.

Extraction of semantically meaningful visual objects from still images and video has enormous applications in video editing, processing, and compression (as in MPEG-4) as well as in search (as in MPEG-7) applications. Extraction of a semantically meaningful object such as a building, a person, a car etc. may be decomposed into extraction of homogeneous regions of the semantic object and performing a "union" of these portions at a later stage. The homogeneity may be in color, texture, or motion. As an example, extraction of a car is considered as extraction of tires, windows and other glass portions, and the body of the car itself.

What is desired is a process that may be used to extract a homogenous color portion of an object.

BRIEF SUMMARY OF THE INVENTION

Accordingly the present invention provides a process for extracting regions of homogeneous color in a digital picture based on a color gradient field with two methods for computing the gradient field -- a weighted Euclidean distance between moment-based feature vectors and a so-called pmf-based distance metric. The digital picture is divided into blocks, and a feature vector is generated for each block as the set of moments for the data in the block. The maximum distance between each block and its nearest neighbors is determined, using either the weighted Euclidean distance metric or the probability mass function-based distance metric, to generate a gradient value for each block. The set of gradient values define the color gradient field. The gradient field is digitized and smoothed, and then segmented into regions of similar color characteristics using a watershed algorithm.

The objects, advantages and other novel features of the present invention are apparent from the following detailed description when read in conjunction with the appended claims and attached drawing.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

Fig. 1 is a block diagram view of an overall process according to the present invention.

Fig. 2 is an illustrative view of an original image.

Fig. 3 is an illustrative view of a segmentation map of the image of Fig. 2 according to a first embodiment of the present invention.

Fig. 4 is an illustrative view of a segmentation map of the image of Fig. 2 according to a second embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The process described here is block-based, i.e. the digital picture is first divided into many non-overlapping rectangular blocks (in general blocks of other shapes and of different sizes, and use of overlapping blocks may be used), and then spatially adjacent blocks that have similar color properties are merged together. This results in the classification of the picture into several spatially contiguous groups of blocks, each group being homogenous in color.

First, segment a digital picture based on a color gradient field, and then use one of two methods for computing that gradient field. The first method makes use of the weighted Euclidean distance between moment-based feature vectors. The second method makes use of the so-called pmf-based distance metric. The overall process is shown in Fig. 1.

The digital input images are assumed to be in YUV format. If the inputs are in a chrominance sub-sampled format such as 4:2:0, 4:1:1 or 4:2:2, the chrominance data is upsampled to generate 4:4:4 material.

Extract one feature vector for each PxQ block of the input picture. There are two stages in the feature vector generation process. In the first stage, transform the data from the original YUV color co-ordinate system into another co-ordinate system known as CIE -- L*a*b* [see *Fundamentals of*

Digital Image Processing, by Anil K. Jain, Prentice-Hall, Section 3.9]. The latter is known to be a perceptually uniform color system, i.e. the Euclidean distance between two points (or colors) in the CIE -- L*a*b* co-ordinate system corresponds to the perceptual difference between the colors.

The next stage in the feature vector generation process is the calculation of the first N moments of the CIE -- L*a*b* data in each block. Thus, each feature vector has 3N components (N moments in *L*, N moments in *a*, and N moments in *b*). (See the Appendix)

The next stage in the region extraction process is that of gradient extraction. Estimate a block-based gradient field for the input picture (i.e. get one scalar gradient value for each PxQ block of the input picture). The gradient at the (i, j)-th block of the input picture is defined as the maximum of the distances between the block's feature vector $f(i,j)$ and its nearest neighbor's feature vectors. (See Appendix)

(In the maximization, let k and l each vary from –1 to +1, but do not allow $k = l = 0$ simultaneously! Also, along the borders of the image, consider only those neighboring blocks that lie inside the image boundaries). Use one of two types of distance functions.

Other methods to select the gradient value from the above set of distances, for example the minimum, median, etc. May be used. It is necessary to evaluate the performance of the segmentation algorithm when such methods are used.

The distance function is simply the weighted Euclidean distance between two vectors. (See Appendix). In the formula, the weighting factors may be used to account for the differences in scale among the various moments. This metric is very easy to implement. In one implementation, set N = 1, i.e. use only the mean values within each PxQ block, and set the weighting factors to unity (this makes sense, since the CIE -- L*a*b* space is perceptually uniform).

The second choice of the distance metric is a little more involved. Here, the fact is exploited that using the moments of the data within the PxQ block, an approximation to the probability mass function (pmf) of that data may be computed. The pmf essentially describes the distribution of the data to be composed of a mixture of several values, with respective probabilities. The values and the probabilities together constitute the pmf. Compute these values using the moments as described in the Appendix.

Thus, the moment-based feature vector of each PxQ block may be converted into a pmf-based representation. With such a representation, then the distance between two feature vectors may be computed via the distance between the two pmf's. For this, make use of the Kolmogorov-Smirnoff (K-S) test, as described in Section 14.3 of *"Numerical Recipes in C"*, 2nd edition, by W. A. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Cambridge University Press. (Essentially, the distance between two pmf's is the area under the absolute value of the difference between

the two cumulative distribution functions, see the above-mentioned chapter for details).

Though the K-S test is prescribed for pmf's of a single variable, the data is in fact three-dimensional (*L*, *a*, and *b* components). Strictly speaking, it is necessary to compute the joint, three-dimensional pmf, and then compute a distance between two pmf's. This is however a very hard problem to solve, and instead a simplifying assumption is made. Assume that the color data in a PxQ block may be modeled by means of three independent pmf's, one each for the *L*, *a*, and *b* components. (See Appendix)

The gradient field, as computed above, yields values that lie along the positive real axis (i.e. can vary from zero to infinity). In practice, the gradient values occupy a finite range, say from minimum to maximum. Digitize the gradient field at a precision of B bits, by dividing the above range into 2^B levels. In one implementation, choose B = 8.

After the gradient field has been digitized, perform morphological preprocessing. This process removes small bumps in the gradient field, and helps the subsequent watershed algorithm to perform a better segmentation. The preprocessing algorithm used has been taken from "*Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking*", by Demin Wang, pages 539 through 546, IEEE Transactions on Circuits and Systems for Video Technology, Volume 8, Number 5, September 1998. "Reconstruction By Erosion" is used as described in

"*Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms*", by Luc Vincent, pages 176 through 201, IEEE Transactions on Image Processing, Volume 2, Issue 2, April 1993. In this process, a smoothing threshold that is 0.7% of the dynamic range of the gradient field is used.

The digitized gradient field, after the above preprocessing, is segmented by what is known as the watershed algorithm. The algorithm description is in the above-mentioned journal article by Luc Vincent. The watershed algorithm divides the gradient field into a set of spatially connected regions, each of which is "smooth" in its interior. Thus, these regions are characterized by having strong gradients at their boundaries. Since the gradient value is proportional to the perceptual difference in color, by the above way of calculating the distance metric, the image is segmented into regions of homogenous color.

Once the input digital image has been segmented into regions that are homogenous in color and are spatially connected, this information may be used in database/search applications. Each region may be represented by one feature vector, consisting of either the same N moments that were used in the segmentation process, or consisting of the pmf-based representation that are computed from those moments. The latter representation is more powerful, because capturing the probability distribution of the data is known to be very useful for indexing visual objects for search applications. In this case the work by Szego

("Orthogonal Polynomials", 4^{th} edition, American Math. Society, Providence, Volume 23, 1975) is used to compute the pmf-based representation from the moments. Then, create an entry for this image in the database, consisting of the classification map together with the characteristic feature vector for each class (region). The use of such an index for database applications is described in a co-pending provisional U.S. Patent Application Serial No.60/118,.

Although in the described implementation non-overlapping rectangular blocks are used, this process may be generalized to blocks of other shapes (square, hexagonal, etc.). Also overlapping blocks may be used, which helps in obtaining a segmentation map that is of higher resolution (than the current block-based segmentation map).

One particular computation of local activity measures has been described, where the moments are computed over rectangular (PxQ) blocks. Activity measures other than moments may be used. Also different block sizes for different areas of the image may be used.

The described pmf-based distance metric uses only two representative values and their probabilities. This metric may be extended by using more representative values (resulting in a more accurate representation of the true probability distribution of the data). A closed form solution for computing more representative values and their corresponding probabilities can be found in the work by Szego.

Other methods than the watershed algorithm may be used to merge blocks. K-means clustering, quadtree segmentation, etc. are possible alternatives.

Thus the present invention provides a process for extracting regions of homogeneous color in a digital picture by segmenting the picture based on a color gradient field, computing the gradient field by one of two distance metrics, digitizing and preprocessing the gradient field, and then segmenting the preprocessed digitized color gradient field with a watershed algorithm.

09195069 .020400